

NO-A106 102

ESTIMATING SYSTEM RELIABILITY: MONTE CARLO METHODS
SENSITIVITY AND ERRORS. (U) NORTH CAROLINA UNIV AT
CHAPEL HILL CURRICULUM IN OPERATIONS R. G S FISHMAN
JAN 87 UNC/ORSA/TR-87/1 AFOSR-TR-87-1092 F/G 12/4

1/1

UNCLASSIFIED

NL





NATIONAL BUREAU OF STANDARDS 1963-4

AD-A186 182

AFOSR-TR- 87 - 1092

OPERATIONS RESEARCH AND SYSTEMS ANALYSIS

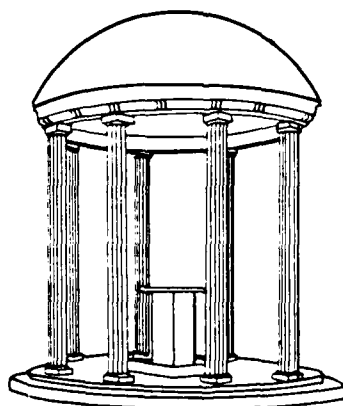
Estimating System Reliability:
Monte Carlo Methods, Sensitivity and
Errors in Input Parameters

George S. Fishman

Technical Report No. UNC/ORSA/TR-87/1

January 1987

UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL



DTIC
ELECTE
OCT 07 1987
S D
E

This document has been approved
for public release and sale; its
distribution is unlimited.

Estimating System Reliability:
Monte Carlo Methods, Sensitivity and
Errors in Input Parameters

George S. Fishman

Technical Report No. UNC/ORSA/TR-87/1

January 1987

Curriculum in Operations Research and
Systems Analysis
University of North Carolina at Chapel Hill

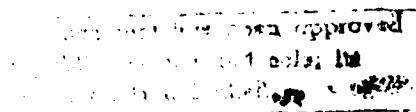


Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC
ELECTE
S OCT 07 1987 D
E

Prepared for invited presentation at the Second International
Workshop on Applied Mathematics and Performance Reliability
Models of Computer/Communication Systems, University of Rome,
Italy, May 25-29, 1987.

This research was supported by the Air Force Office of Scientific
Research under grant AFOSR-84-0140. Reproduction in whole or
part is permitted for any purpose of the United States
Government.



Abstract

In assessing the reliability of a binary-state system whose components fail randomly and independently, it is often of interest to compute the probability that two terminals or nodes s and t are connected and, more generally, the probability that all n nodes of a system are connected. In the case of multi-state flow systems whose components have randomly determined capacity flow levels, interest often focuses on the distribution of maximal flow from a source node s to a terminal node t . Both the binary-state connectivity model and the multi-state flow model can be used to represent communication and transportation networks.

Since computing these probabilities belongs to the class of #P-hard problems, no algorithms are known to solve them exactly in time polynomial in the input. As an alternative method of solution, Monte Carlo sampling plans are being developed to estimate these quantities, especially for large systems. This method has the advantage that errors in evaluation are essentially independent of the size of the system. This paper describes how these Monte Carlo sampling plans are designed to use prior information on the system under study to obtain estimates with specified accuracy at considerably less cost than crude Monte Carlo sampling would require. Such plans are called variance reducing techniques and represent an integral component of the Monte Carlo method. Also, this prior information about system design enables one to estimate, prior to performing any sampling, the maximal sample size required to achieve a specified accuracy when using the variance reducing techniques. These worst case bounds on sample size are especially helpful when a limited computing budget is available.

The paper also describes how Monte Carlo sampling data accumulated to estimate the aforementioned reliability and flow probabilities can be used to perform sensitivity analyses with

only incidental additional cost. These analyses include estimating the effect of improving component reliability in the binary-state case and the effect of increasing the probability that a component operates at full capacity in the multi-state flow case.

Virtually all computations of system reliability measures assume that component reliability probabilities, which are the input to the computation, are known with certainty. This is rarely the case and the errors that do exist in these input data directly affect the quality of the reliability computation, whether it be an exact or a Monte Carlo based calculation. This issue is discussed in detail. In particular, the paper describes the error-induced bias and variance as functions of the parameters of the system under study and shows how these measures can adversely affect the interpretation of a reliability calculation.

Key Words: maximal flow; Monte Carlo sampling; reliability;
s-t connectedness

Introduction

The computation of system reliability from component reliabilities presents a host of non-trivial problems for systems of varying sizes. These include the functional relationship between the time required to compute system reliability and system size. A second problem concerns how system reliability varies as component reliabilities vary. A third problem concerns how statistical errors in estimating component reliabilities affect the accuracy of the system reliability computation.

This paper describes Monte Carlo techniques which provide useful answers to the first two problems and presents an analysis which establishes the potential seriousness of the third problem in practice. The presentation summarizes research reported in [3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. To put the results in perspective, we begin with a description of the system reliability concepts that are of interest to us.

1. Characterizations of System Reliability

Consider the network $G = (\underline{V}, \underline{E})$ with node set \underline{V} and arc set \underline{E} . For convenience of exposition, assume that nodes represent components that function perfectly and that arcs represent components that fail randomly and independently. To characterize G more completely, we define:

- r = number of distinct types of components
- $F_i(\tau)$ = probability that a component of type i functions at time $\tau \geq 0$ for $i=1, \dots, r$.
- $\underline{F}(\tau) = (F_1(\tau), \dots, F_r(\tau))$
- \underline{E}_i = set of arcs that use components of type i
 $(\underline{E}_i \cap \underline{E}_j = \emptyset \quad i \neq j \quad \underline{E} = \bigcup_{i=1}^r \underline{E}_i)$
- k_i = $|\underline{E}_i|$ number of components of type i
- $\underline{k} = (k_1, \dots, k_r)$
- e_{ij} = j th arc in \underline{E}_i
- x_{ij} = 1 if arc e_{ij} functions, = 0 otherwise
- $x_i = \sum_{j=1}^{k_i} x_{ij}$ = number of arcs of type i that function

$$\underline{x} = (x_{11}, \dots, x_{1k_1}; x_{21}, \dots, x_{2k_2}; \dots; x_{r1}, \dots, x_{rk_r})$$

\underline{X} = set of all arc states \underline{x}

$$P(\underline{x}, \underline{k}, \underline{F}(\tau)) = \prod_{i=1}^r [F_i(\tau)]^{x_i} [1-F_i(\tau)]^{k_i-x_i} \quad \underline{x} \in \underline{X} \quad (1)$$

= probability mass function of states in \underline{X} at time τ

$$\phi(\underline{x}) = 1 \text{ if the system functions, } = 0 \text{ otherwise} \quad (2)$$

$$g(\underline{F}(\tau)) = \sum_{\underline{x} \in \underline{X}} \phi(\underline{x}) P(\underline{x}, \underline{k}, \underline{F}(\tau)) = \text{probability that the system functions at time } \tau. \quad (3)$$

The interpretation of the system reliability $g(\underline{F}(\tau))$ varies with the type of system under study. For a binary-state system with binary-state components, we consider the \underline{T} -connectedness problem. Let \underline{T} denote a subset of \underline{V} and let

$$\begin{aligned} \phi(\underline{x}) &= 1 \text{ if all nodes in } \underline{T} \text{ are connected when state } \underline{x} \text{ occurs} \\ &= 0 \text{ otherwise.} \end{aligned}$$

Then

$$g(\underline{F}(\tau)) = \text{probability that all nodes in } \underline{T} \text{ are connected at time } \tau. \quad (4)$$

When $\underline{T} = \{s, t\} \subset \underline{V}$, this is called the s-t connectedness problem. When $\underline{T} = \underline{V}$, it is called the all-terminal connectedness problem. These representations are useful in studying the vulnerability of communication systems.

One can also use the network representation to study the reliability of multi-state systems with binary-state components. Suppose G is a directed acyclic flow network with source node s and terminal node t and that at time τ each component of type i has flow capacities zero with probability $1-F_i(\tau)$ and $b_i > 0$ with probability $F_i(\tau)$. Let

$g(F(\tau))$ at all $\tau \in \tau_1 = \{\tau_1, \dots, \tau_k\}$ takes $O(k)$ times as long as computing $g(F(\tau))$ at a single point.

To overcome this limitation to exact computation, one can resort to a Monte Carlo sampling experiment wherein one approximates $g(F(\tau))$ by an estimate whose error of approximation decreases as the number of independent trials, on which the estimate is based, increases. Many alternative sampling plans exist and, in principle, one prefers a plan that exploits prior knowledge about the system of interest to achieve an error of specified size at less cost than alternative methods allow.

We now describe such a sampling plan that applies with small modifications to the estimation of diverse reliability measures such as (4), (5), (6), (7) and (8), when they are appropriate. For conciseness of exposition, we assume that estimation is to be performed for fixed $F(\tau)$ and k and hereafter suppress these arguments unless they are needed to avoid ambiguity. Reference [4] provides a comparison of some of the Monte Carlo techniques to be described here with the proposed Monte Carlo sampling plans of other writers.

3. Using Prior Information

Assume that sufficient prior information exists about the system under study to enable one to identify two binary functions $\{\phi_1(x) \mid x \in X\}$ and $\{\phi_2(x) \mid x \in X\}$ with the property

$$\phi_1(x) \leq \phi(x) \leq \phi_2(x) \quad (9)$$

where $\phi(x)$ is defined in (2). Let

$$g_1 = \sum_{x \in X} \phi_1(x) P(x)$$

so that

$$g_1 \leq g \leq g_2. \quad (10)$$

Now let

$$Q(\underline{x}) = \frac{\phi_2(\underline{x}) - \phi_1(\underline{x})}{g_2 - g_1} P(\underline{x}) \quad \underline{x} \in X, \quad (11)$$

which is a probability mass function. Suppose that one can compute g_1 and g_2 exactly and at reasonable cost as a function of the size of G . Then the ordering relationships (9) and (10) together with the probability mass function $\{Q(\underline{x}) \mid \underline{x} \in X\}$ in (11) allow one to benefit from two alternative modes of sampling.

3.1 Importance Sampling [3,6,7,8,10]

Here one concentrates sampling in the region of the state space $X_{12} = \{\underline{x} \in X \mid \phi_1(\underline{x})=0 \text{ and } \phi_2(\underline{x})=1\}$ as follows:

1. Set $S \leftarrow 0$.
2. On each of K independent trials:
 - 2a. Sample \underline{x} from $\{Q(\underline{x}) \mid \underline{x} \in X\}$.
 - 2b. Compute $\phi(\underline{x})$.
 - 2c. Set $S \leftarrow S + \phi(\underline{x})$.
3. Compute summary statistics
 - 3a. $\hat{g}_K = g_1 + (g_2 - g_1)S/K$.
 - 3b. $V(\hat{g}_K) = (g_2 - \hat{g}_K)(\hat{g}_K - g_1)/(K-1)$.

Here \hat{g}_K is an unbiased estimator of g with

$$\text{var } \hat{g}_K = (g_2 - g)(g - g_1)/K \quad (12)$$

and $V(\hat{g}_K)$ is an unbiased estimator of $\text{var } \hat{g}_K$.

One way to assess the benefit of this sampling plan is to compare it with the results for a crude Monte Carlo sampling estimate \bar{g}_K of g , using $\{P(\underline{x})\}$. This too is unbiased but with

$$\text{var } \bar{g}_K = g(1-g)/K. \quad (13)$$

The ratio of variances is then

$$\begin{aligned}
 R_I &= \text{var } \bar{g}_K / \text{var } \hat{g}_K \\
 &\geq 1 / \{g_2 - g_1 - 2[g_1 g_2 (1 - g_1)(1 - g_2)]^{1/2}\} \\
 &\geq 1.
 \end{aligned}
 \tag{14}$$

Since the lower bound (14) is computable prior to sampling, one can obtain a clear understanding of the least favorable result that the sampling plan can induce before employing it.

The variance ratio merely measures one aspect of the importance sampling plan. Let $c(P)$ and $c(Q)$ denote the mean times per replication of employing sampling plans based on $\{P(x)\}$ in (1) and $\{Q(x)\}$ in (11) respectively. Note that each of these includes the cost of computing $\phi(x)$. Then

$$W_I = \frac{c(P)}{c(Q)} R_I$$

denotes the mean time required to achieve the same variance with crude Monte Carlo sampling as one obtains in one unit of time with the importance sampling plan. For importance sampling to be beneficial, one requires $W_I > 1$, preferably considerably greater than 1. Reference [3] describes how to construct $\{Q(x)\}$ and how to sample from it for \underline{I} -connectivity, and [10] describes the corresponding construction for the reliability measure $g(F(\tau), z)$ in (5). In both cases, the ability to determine sets of edge-disjoint minimal spanning trees and minimal cutsets enables one to identify the binary functions $\{\phi_1(x)\}$ and $\phi_2(x)\}$ needed to achieve the variance reduction. References [3] and [10] also describe how to compute confidence intervals for g that hold for finite K . These considerably improve on normal confidence intervals whose use inevitably introduces an additional error of approximation.

3.2 Control Variates [9]

This plan uses inequalities (9) and (10), but samples the state vector \underline{x} from $\{P(\underline{x})\}$ in (1) instead of from $\{Q(\underline{x})\}$ in (11). In particular:

1. Set $S \leftarrow 0$, $S_1 \leftarrow 0$ and $S_2 \leftarrow 0$.
2. On each of K trials:
 - 2a. Sample \underline{x} from $\{P(\underline{x})\}$.
 - 2b. Compute $\phi_1(\underline{x})$, $\phi(\underline{x})$ and $\phi_2(\underline{x})$.
 - 2c. Set $S_1 \leftarrow S_1 + \phi_1(\underline{x})$, $S \leftarrow S + \phi(\underline{x})$ and $S_2 \leftarrow S_2 + \phi_2(\underline{x})$.
3. Compute summary statistic

$$\bar{g}_K = \frac{S(g_2 - g_1) + (S/K - g_2)(S_1 - Kg_1) + (g_1 - S/K)(S_2 - Kg_2) - g_1}{(K+1)(g_2 - g_1) - 1}. \quad (15)$$

The resulting statistic is an unbiased estimator of g with

$$\lim_{K \rightarrow \infty} K \text{ var } \bar{g}_K = (g_2 - g)(g - g_1)/(g_2 - g_1). \quad (16)$$

An unbiased estimator of $\text{var } \bar{g}_K$ is also given in [10].

Observe that for large K

$$\text{var } \bar{g}_K = \text{var } \hat{g}_K / (g_2 - g_1) > \text{var } \hat{g}_K. \quad (17)$$

Although sampling from $\{Q(\underline{x})\}$ for \hat{g}_K consumes more time than sampling from $\{P(\underline{x})\}$ for \bar{g}_K does, experience has shown that importance sampling tends to be more statistically efficient than the method of control variates is for reliability estimation at a single point. However, in the network flow model the control variate approach produces sample data that allow estimation at more than one flow point z whereas importance sampling does not. In particular, [10] shows how this can be done in estimating the complementary distribution of maximal s - t flow $\{g(F(\tau), z)\}$ at a set of points $z \in \underline{Z}$. Confidence intervals for $\{g(F(\tau), z) \mid z \in \underline{Z}\}$ are also given there.

4. Determining the Worst Case Sample Size

In addition to producing a statistically more efficient estimator than crude Monte Carlo sampling does, an effective Monte Carlo technique also provides information that enables one to determine how large a sample size is necessary to achieve an estimate with a specified upper bound on statistical error. Both the importance sampling and control variate approaches provide such a result based on a worst case scenario.

Let X_1, \dots, X_K denote i.i.d. random variables with $\mu = EX_i$ and $\text{pr}(a \leq X_i \leq b) = 1$ for all $i=1, \dots, K$. Then for $\lambda \equiv (\mu-a)/(b-a)$ and $\epsilon < \min(\lambda, 1-\lambda)$, Hoeffding (1963) shows that

$$\text{pr}(\bar{X}_K - \mu \geq \epsilon) \leq R^K(\lambda, \epsilon/(b-a))$$

and

$$\text{pr}(-\bar{X}_K + \mu \geq \epsilon/(b-a)) \leq R^K(1-\lambda, \epsilon/(b-a))$$

where

$$R(\lambda, \theta) = \left(\frac{\lambda}{\lambda+\theta}\right)^{\lambda+\theta} \left(\frac{1-\lambda}{1-\lambda-\theta}\right)^{1-\lambda-\theta} \quad 0 < \theta < \min(\lambda, 1-\lambda), \quad 0 \leq \lambda \leq 1.$$

Note that R has a unique minimum for λ in $[0, 1-\epsilon/(b-a)]$.

Suppose that one wants to determine the minimal sample size K^* such that

$$\text{pr}(|\bar{X}_{K^*} - \mu| \leq \epsilon) \geq 1 - \alpha \quad (18)$$

for sampling absolute accuracy ϵ and confidence level $1 - \alpha$. Then collecting

$$K^* = \ln(\alpha/2) / \max_{0 \leq \lambda \leq 1} \ln R(\lambda, \epsilon/(b-a)) \quad (19)$$

observations assures the absolute error bounds in (18).

Now suppose that $\mu > 0$ and consider the relative error criterion

$$\text{pr}(|\bar{X}_{K^*} - \mu| \leq \epsilon \mu) \geq 1 - \alpha. \quad (20)$$

Then K^* is again computable as in (19) with $\epsilon \mu$ replacing ϵ so that

$$K^* = \ln(\alpha/2) / \min \left[\max_{0 \leq \lambda \leq 1} \ln R(\lambda, (\frac{a}{b-a} + \lambda)\epsilon), \max_{0 \leq \lambda \leq 1} \ln R(1-\lambda, (\frac{a}{b-a} + \lambda)\epsilon) \right]. \quad (21)$$

For the reliability problem using importance sampling with $\bar{X}_K = \hat{g}_K$, $a = g_1$ and $b = g_2$, the significance of (19) and (21) lies in the realization that one can simply take these numbers of observations and be guaranteed the corresponding specified accuracy without any subsequent analysis after computation of \hat{g}_K .

A worst case analysis also exists for the control variate sampling plan. Let

$$r(\lambda, \epsilon, h, g_1, g_2) = e^{-h\epsilon} \{1 - g_2 + g_1 + (g_2 - g_1)[\lambda e^{(1-\lambda)h} + (1-\lambda)e^{-\lambda h}]\} \quad (22)$$

$$\lambda \equiv \frac{g - g_1}{g_2 - g_1}, \quad h \geq 0 \quad 0 < \epsilon \leq \min(\lambda, 1-\lambda).$$

To ensure an absolute accuracy ϵ with probability $\geq 1 - \alpha$, one needs no more than

$$K^* = \ln(\alpha/2) / \min \left\{ \max_{0 \leq \lambda \leq 1} [\ln \min_{h \geq 0} R(\lambda, h, g_1, g_2)], \max_{0 \leq \lambda \leq 1} [\ln \min_{h \geq 0} R(1-\lambda, h, g_1, g_2)] \right\} \quad (23)$$

observations. In spite of its formidable appearance, one can evaluate (23) using a modification of algorithm A in [9].

5. Sensitivity [12]

In addition to estimating the reliability $g(F(\tau))$ at a fixed time τ , the Monte Carlo method provides a way, at small marginal cost, of estimating the more general reliability function $\{g(\underline{q}) \mid \underline{q} = (q_1, \dots, q_r) \in \underline{W}\}$ where q_i = probability that a component of type i functions $i=1, \dots, r$ and \underline{W} is a set of component reliability vectors of interest. Here q_i for component i may be a function of time (i.e. $q_i = F_i(\tau)$) or it may reflect potential component reliability improvements the effect of which on system reliability is

of interest.

With this change of nomenclature, one has

$$P(\underline{x}, \underline{k}, \underline{q}) = \prod_{i=1}^r q_i^{x_i} (1-q_i)^{k_i - x_i} \quad (24)$$

and

$$g(\underline{q}) = \sum_{\underline{x} \in X} \phi(\underline{x}) P(\underline{x}, \underline{k}, \underline{q}) \quad \underline{q} \in W. \quad (25)$$

Suppose that one elects to perform K independent replications with component reliability \underline{p} using $\{Q(\underline{x}, \underline{k}, \underline{p})\}$ (importance sampling). Let $\underline{x}^{(j)}$ denote the sample component state vector on replication j. The estimators

$$\hat{g}_{1K}(\underline{q}) = g_1(\underline{q}) + [g_2(\underline{p}) - g_1(\underline{p})] K^{-1} \sum_{j=1}^K \phi(\underline{x}^{(j)}) \frac{P(\underline{x}^{(j)}, \underline{k}, \underline{q})}{P(\underline{x}^{(j)}, \underline{k}, \underline{p})} \quad (26)$$

and

$$\hat{g}_{2K}(\underline{q}) = g_2(\underline{q}) - [g_2(\underline{p}) - g_1(\underline{p})] K^{-1} \sum_{j=1}^K [1 - \phi(\underline{x}^{(j)})] \frac{P(\underline{x}^{(j)}, \underline{k}, \underline{q})}{P(\underline{x}^{(j)}, \underline{k}, \underline{p})} \quad (27)$$

are unbiased estimators of $g(\underline{q})$ with

$$Kv_1(\underline{q}) = K \text{ var } \hat{g}_{1K}(\underline{q}) = \{c[g_2(\underline{p}) - g_1(\underline{p})][g(\underline{q}^*) - g_1(\underline{q}^*)] - [g(\underline{q}) - g_1(\underline{q})][g_2(\underline{q}) - g_1(\underline{q})]\} \\ + [g_2(\underline{q}) - g(\underline{q})][g(\underline{q}) - g_1(\underline{q})] \quad (28)$$

$$Kv_2(\underline{q}) = K \text{ var } \hat{g}_{2K}(\underline{q}) = \{c[g_2(\underline{p}) - g_1(\underline{p})][g_2(\underline{q}^*) - g(\underline{q}^*)] - [g_2(\underline{q}) - g(\underline{q})][g_2(\underline{q}) - g_1(\underline{q})]\} \\ + [g_2(\underline{q}) - g(\underline{q})][g(\underline{q}) - g_1(\underline{q})] \quad (29)$$

$$Kc_{12}(\underline{q}) = K \text{ cov}[\hat{g}_{1K}(\underline{q}), \hat{g}_{2K}(\underline{q})] = [g_2(\underline{q}) - g(\underline{q})][g(\underline{q}) - g_1(\underline{q})] \quad (30)$$

where

$$c = \prod_{i=1}^r [q_i^2/p_i + (1-q_i)^2/(1-p_i)]^{k_i},$$

$$\underline{q}^* = (q_1^*, \dots, q_r^*)$$

and

$$q_i^* = q_i^2/cp_i \quad i=1, \dots, r.$$

In principle, (26) and (27) enable one to estimate $g(\underline{q})$ for all $\underline{q} \in W$ from the single set of replications obtained by sampling with component reliabilities \underline{p} . Since $[g_2(\underline{q}) - g(\underline{q})][g(\underline{q}) - g_1(\underline{q})]$ in (28) and (29) is the variance of the point estimator $\hat{g}_K(\underline{q})$ based on sampling with component reliabilities \underline{q} , the quantities in curly brackets are the incremental changes in variances that result from sampling with \underline{p} instead of \underline{q} . Of most importance, it is entirely possible for one of these quantities to be negative, implying that an estimate of $g(\underline{q})$ with smaller variance than $\hat{g}_K(\underline{q})$ is possible.

To put this last observation in perspective, consider the estimator

$$\bar{g}_K(\underline{q}) = \theta \hat{g}_{1K}(\underline{q}) + (1-\theta) \hat{g}_{2K}(\underline{q}) \quad 0 \leq \theta \leq 1. \quad (31)$$

which achieves minimal variance by choosing

$$\begin{aligned} \theta &= 1 && \text{if } v_1(\underline{q}) \leq c_{12}(\underline{q}) \\ &= \frac{v_2(\underline{q}) - c_{12}(\underline{q})}{v_1(\underline{q}) + v_2(\underline{q}) - 2c_{12}(\underline{q})} && v_1(\underline{q}), v_2(\underline{q}) \geq c_{12}(\underline{q}) \\ &= 0 && v_2(\underline{q}) \leq c_{12}(\underline{q}). \end{aligned} \quad (32)$$

Observe that $v_1(\underline{q}) \leq c_{12}(\underline{q})$ implies that $\hat{g}_{1K}(\underline{q})$ in (26) has smaller variance than $\hat{g}_K(\underline{q})$ in (12) whereas $v_2(\underline{q}) \leq c_{12}(\underline{q})$ implies that $\hat{g}_{2K}(\underline{q})$ in (27) has smaller variance than $\hat{g}_K(\underline{q})$ in (12).

Experience with this method of estimation indicates that for moderately high component reliabilities $v_2(\underline{q}) \leq c_{12}(\underline{q})$ is often satisfied with substantial reductions in variance at each \underline{q} in \underline{W} , as compared to the results from importance sampling at a point in Section 3.1. Moreover, all these estimates follow from just one sampling experiment at \underline{p} . The development of a confidence region for the reliability function $\{g(\underline{q}) \mid \underline{q} \in \underline{W}\}$ is now underway, as is a study to determine the optimal \underline{p} at which to sample from the component state space \underline{X} .

6. Errors in Input Parameters [11]

Regardless of whether an exact or Monte Carlo method is used in computing system reliability $g(\underline{q})$, it is customary to assume that the numerical values of the component reliabilities $\underline{q} = (q_1, \dots, q_r)$ are exact. In reality they are not, and, as we now show that ignoring this potential source of error can give a misleading interpretation to the final numerical value computed for $g(\underline{q})$.

Suppose one tests n_i components of type i for $i=1, \dots, r$. Each test begins with a new component functioning. Let Z_{ij} denote the outcome of the j th test of component of type i where $Z_{ij} = 1$ if the component functions at the end of the test period and $Z_{ij} = 0$ if the component fails prior to the end of the test period. Presumably each component of type i is tested under identical conditions that resemble the system environment. Then one has the data vectors $\underline{Z}_i = \{Z_{i1}, \dots, Z_{in_i}\}$ for $i=1, \dots, r$ where the elements of \underline{Z}_i are independent and identically distributed with $q_i = E Z_{ij} \quad j=1, \dots, n_i$ and $\underline{Z}_1, \dots, \underline{Z}_r$ are independent. Also,

$$\hat{q}_i = n_i^{-1} \sum_{j=1}^{n_i} Z_{ij} \quad (33)$$

is the maximum likelihood estimator of q_i with

$$q_i = E\hat{q}_i$$

$$\text{var } q_i = q_i(1-q_i)/K$$

$$E(\hat{q}_i - q_i)^k = O(1/n_i^{\lfloor (k+1)/2 \rfloor}) \quad k \geq 2 \quad \text{as } n_i \rightarrow \infty.$$

Let $\hat{\underline{q}} = (\hat{q}_1, \dots, \hat{q}_r)$. Then it is not unusual to compute system reliability as $g(\hat{\underline{q}})$ with no mention of the sampling error that replacing \underline{q} by $\hat{\underline{q}}$ induces. Unfortunately, the results in Fishman (1987) show that

$$Eg(\hat{\underline{q}}) - g(\underline{q}) \sim w(\underline{k}, \underline{q}, \underline{n}) \quad \text{as } \min_{1 \leq i \leq r} n_i \rightarrow \infty \quad (34)$$

and

$$\text{var } g(\hat{\underline{q}}) \sim v(\underline{k}, \underline{q}, \underline{n}) \quad \text{as } \min_{1 \leq i \leq r} n_i \rightarrow \infty \quad (35)$$

where

$$\underline{n} = (n_1, \dots, n_r),$$

$$w(\underline{k}, \underline{q}, \underline{n}) = \sum_{\underline{x} \in \underline{X}} \left\{ \sum_{\underline{x} \in \underline{X}} \phi(\underline{x}) P(\underline{x}, \underline{k}, \underline{q}) \left[\frac{k_i(k_i-1)q_i^2 - 2x_i(k_i-1)q_i + x_i(x_i-1)}{q_i^2(1-q_i)^2} \right] \right\} \frac{q_i(1-q_i)}{2n_i} \quad (36)$$

and

$$v(\underline{k}, \underline{q}, \underline{n}) = \sum_{i=1}^r \left\{ \sum_{\underline{x} \in \underline{X}} \phi(\underline{x}) P(\underline{x}, \underline{k}, \underline{q}) \left[\frac{x_i - k_i q_i}{q_i(1-q_i)} \right] \right\}^2 \frac{q_i(1-q_i)}{n_i}. \quad (37)$$

Expressions (34) and (35) imply that even in the case of an exact computation (no Monte Carlo sampling), a substantial degree of statistical error can arise. Clearly $g(\hat{\underline{q}})$ is biased. Moreover, this bias (36) and the variance (37) grow quadratically with k_1, \dots, k_r the number of components of each type used in the system and linearly with the number r of different types of components.

Of special interest is the observation that if all components of type i are in series and \hat{q}_i replaces q_i in the reliability computation, then the resulting reliability overstates the true reliability. Conversely, an understatement of reliability occurs when all components of type i are in parallel.

References

1. Ball, M.O. (1979). Computing network reliability, Oper. Res., 27, 823-838
2. Buzacott, J.A. (1983). A recursive algorithm for directed-graph reliability, Networks, 13, 241-246.
3. Fishman, George S. (1986). A Monte Carlo sampling plan for estimating network reliability, Oper. Res., 34, 581-594.
4. Fishman, George S. (1986). A comparison of four Monte Carlo methods for estimating the probability of s-t connectedness, IEEE Trans. on Rel., 35, 145-154.
5. Fishman, George S. (1986). A Monte Carlo sampling plan for estimating reliability parameters and related functions, to appear in Networks.
6. Fishman, George S. (1986). The distribution of maximum flow with applications to multi-state reliability systems, to appear in Oper. Res.
7. Fishman, George S. (1986). Maximum flow and critical cutset as descriptors of multi-state systems with randomly capacitated components, to appear in Computers and Operations Research.
8. Fishman, George S. (1986). Estimating the s-t reliability function using importance and stratified sampling, Technical Report No. UNC/ORSA/TR-86/5, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.
9. Fishman, G.S. (1986). Monte Carlo control variates and stochastic ordering, Technical Report No. UNC/ORSA/TR-86/16, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill.
10. Fishman, G.S. (1986). Monte Carlo estimation of the maximal flow distribution with discrete stochastic arc capacity levels, Technical Report No. UNC/ORSA/TR-86/18, Curriculum in Operations Research and Systems Analysis, University of North Carolina.
11. Fishman, G.S. (1987). Adjusting system reliability to account for statistical errors in component reliabilities, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, in preparation.

12. Fishman, George S. (1987). Sensitivity analysis for the estimation of the s-t reliability function, Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, in preparation.
13. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc., 58, 13-29.
14. Valiant, L.G. (1979). The complexity of enumeration and reliability problems, SIAM J. Comp., 8, 410-421.

ADA186182

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY N/A		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina at Chapel Hill		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) 210 Smith Building 128A Chapel Hill, NC 27514		7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332-6448	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0140
8c. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, Washington, DC 20332-6448		10. SOURCE OF FUNDING NOS.	
		PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304
11. TITLE (Include Security Classification) Estimating system reliability: Monte Carlo methods, sensitivity and errors in input parameters		12. PERSONAL AUTHOR(S) George S. Fishman	
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM TO	14. DATE OF REPORT (Yr., Mo., Day) January 1987	15. PAGE COUNT 19
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>In assessing the reliability of a binary-state system whose components fail randomly and independently, it is often of interest to compute the probability that two terminals or nodes s and t are connected and, more generally, the probability that all n nodes of a system are connected. In the case of multi-state flow systems whose components have randomly determined capacity flow levels, interest often focuses on the distribution of maximal flow from a source node s to a terminal node t. Both the binary-state connectivity model and the multi-state flow model can be used to represent communication and transportation networks.</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL George S. Fishman		22b. TELEPHONE NUMBER (Include Area Code) 919-962-8401	22c. OFFICE SYMBOL AFOSR/NM

-2-

Since computing these probabilities belongs to the class of #P-hard problems, no algorithms are known to solve them exactly in time polynomial in the input. As an alternative method of solution, Monte Carlo sampling plans are being developed to estimate these quantities, especially for large systems. This method has the advantage that errors in evaluation are essentially independent of the size of the system. This paper describes how these Monte Carlo sampling plans are designed to use prior information on the system under study to obtain estimates with specified accuracy at considerably less cost than crude Monte Carlo sampling would require. Such plans are called variance reducing techniques and represent an integral component of the Monte Carlo method. Also, this prior information about system design enables one to estimate, prior to performing any sampling, the maximal sample size required to achieve a specified accuracy when using the variance reducing techniques. These worst case bounds on sample size are especially helpful when a limited computing budget is available.

The paper also describes how Monte Carlo sampling data accumulated to estimate the aforementioned reliability and flow probabilities can be used to perform sensitivity analyses with only incidental additional cost. These analyses include estimating the effect of improving component reliability in the binary state case and the effect of increasing the probability that a component operates at full capacity in the multi-state flow case.

Virtually all computations of system reliability measures assume that component reliability probabilities, which are the input to the computation, are known with certainty. This is rarely the case and the errors that do exist in these input data directly affect the quality of the reliability computation, whether it be an exact or a Monte Carlo based calculation. This issue is discussed in detail. In particular, the paper describes the error-induced bias and variance as functions of the parameters of the system under study and shows how these measures can adversely affect the interpretation of a reliability calculation.

END
DATE
FILMED
JAN
1988